DOCUMENT RESUME

ED 156 217 IR 006 450

TITLE	The Education Division's Joint Dissemination Review Panel: Three Papers. Product 2.
INSTITUTION	Educational Products Information Exchange Inst., New York, N.Y.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	1 Mar 77
CONTRACT	400-75-0023
NOTE	34p.; For related documents, see IR 006 451-452
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Decision Making; Educational Development; Educational Resources; Educational Technology; Evaluation Criteria; *Evaluation Methods; *Federal Government; Government Publications; *Information Dissemination; *Instructional Materials; *Selection
IDENTIFIERS	Joint Dissemination Review Panel

ABSTRACT

This document, which is the second product of an EPIE (Educational Products Information Exchange) project investigating the criteria used to select instructional materials, consists of three papers dealing with the Education Division of the JDRP (Joint Dissemination Review Panel). Paper I presents the purposes, procedures, and criteria established by the JDRP to be used in judging the effectiveness of educational products, materials, practices, and methods developed with funds from either OE (Office of Education) or NIE (National Institute of Education). The JDRP ensures that the two agencies apply the same criteria for judging evidence of effectiveness before federal dissemination. Paper II, by Dr. Lois-ellin Datta, explains the JDRP controversy centering around four issues: (1) criteria used, (2) review of proprietary products, (3) equity of opportunity in providing evidence, and (4) impact of evaluation practice and flow of products to the field. Paper III, by G. Kasten Tallmadge, suggests that JDRP approval not be considered a sufficient condition for federal expenditure for dissemination, but that JDRP approval constitute an approval for further evaluation of the product being considered. (Author/STS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED156217

NIE Products on Instructional Materials Selection

Created under National Institute of Education

(D.H.E.W.) Contract #400-75-0023

By EPIE Institute, 475 Riverside Drive, New York, N.Y. 10027

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Product 2. THE EDUCATION DIVISION'S JOINT DISSEMINATION REVIEW PANEL: THREE PAPERS

A. "The Education Division's Joint Dissemination Review Panel (JDRP)" 1

B. "The External Implications of an Internal Review of Effectiveness: The DHEW Education Division's Joint Dissemination Review Panel" - by Lois-ellin Datta 16

C. "Comments on Dr. Lois-ellin Datta's Paper" - by G. Kasten Tallmadge 28

IR006450

A. The Education Division's Joint Dissemination Review Panel (JDRP)

PURPOSE, PROCEDURES, AND CRITERIA

Background and Purpose

The United States Congress and educators expect the federal government to disseminate products, materials, practices, and methods which will improve educational quality in U.S. schools. Numerous pieces of legislation have stressed this responsibility of the Office of Education and the National Institute of Education. Various efforts have been made over the years to discharge it.

Eagerness on the part of federal officials to disseminate exemplary educational methods has not always been matched by assurance that the items disseminated were in fact exemplary. Federal program staffs have disseminated materials which seem innovative, intriguing, or which enjoyed a good reputation in the schools where they were used. Seldom was it determined that such products were supported by objective evidence of effectiveness. State and local educators have been persuaded by federal sponsorship to implement practices which have not produced their promised results.

The federal government is irresponsible if it does not take steps to assure that the educational materials it disseminates are backed by objective evidence of effectiveness. The Education Division thus has established a Joint (OE and NIE) Dissemination Review Panel to review evidence of effectiveness of potential exemplary products and practices developed with funds provided by either agency. The Joint OE/NIE Panel insures that the two agencies apply the same criteria for judging evidence of effectiveness prior to federal dissemination.

The Panel serves as one of the internal quality control mechanisms for the dissemination activities of the Education Division. The Panel is not intended in any way to pass on or constrain the dissemination activities of state and local education agencies. It acts as a self-restraining mechanism by which the Education Division facilitates informed local choice. Through the Panel, when an Education Division program office representing the federal government claims that a particular educational practice or product is exemplary and is worthy of implementation by school systems, that assertion will be substantiated.

Operations

The Panel reviews all products/materials and practices/methods proposed for dissemination by any Education Division program office. Panel review is required before any Education Division program office can approve use of federal funds for the dissemination of any practices or products for which claims of effectiveness are made. In addition, review is required for other no-cost activities involving Education Division endorsement of or advocacy for a product or practice for which claims of effectiveness are made. The review is intended to insure that dissemination of materials/practices are supported by objective evidence of effectiveness.

"Products" and "materials" refer to transportable, tangible or packaged programs or practices such as curriculum materials, educational management systems, computer programs, staff training materials, guides, models, administrative procedures, Project Information Packages (PIPS), etc. "Practices" and "methods" refer to systematic programs, procedures, or techniques whose adoption or adaptation requires assistance from the originators or their representatives.

The Panel does not recommend or review any proposed dissemination strategies. A Panel decision that the evidence of effectiveness is compelling may not be the only review

prior to dissemination. Decisions on how to disseminate Panel-reviewed products or practices are left to the sponsoring agencies and programs in the Education Division. It is possible for one Education Division program to submit and another program to disseminate.

The Panel (and its predecessor unit, the OE Dissemination Review Panel) has existed for three years.* It has met over 60 times, reviewed over 280 submissions, and approved about 55% of them for dissemination. In many cases, items not favorably received at the initial review are resubmitted with additional information supporting their evidence of effectiveness.

Composition

The Panel is composed of 22 members appointed by the Commissioner of Education and the Director of NIE. Members are selected on the basis of their ability to analyze evaluation-based evidence on the effectiveness of products and practices, and on their general experience in education. The Panel meets periodically to review submissions from Education Division programs, but not those from SEA's, LEA's, or other outside organizations.

Submissions

As noted above, the Panel is one device for upgrading the quality of the Education Division's dissemination activities. It is not a self-appointed national approval body. The Form for Submitting Material to the Joint Dissemination Review Panel was developed so that Education Division programs can bring together evidence of the effectiveness of products or practices and present them to the Panel. This form

*This document is designed to update a November 1974 statement on "The Office of Education's Dissemination Review Panel: Procedures and Criteria."

calls for a brief description of each innovation (including cautions relating to its use), assurances and claims about the innovation, information on how the innovation was tested, and a summary of the evidence supporting claims of effectiveness. This material is verified for factual accuracy by the submitting federal program before review by the Panel.

Procedures for Submission and Review

A. Education Division Agency or Program Responsibilities Prior to Panel Review

1. Identify potential products or practices* which the federal program may want to state or imply are effective, exemplary, promising, or worthy of dissemination support. Panel review is not required for catalogs and purely descriptive lists where there is no claim of effectiveness made and no recommendation for adoption.

2. Prereview the claims of effectiveness for factual accuracy. Prereview for social fairness and with respect to possible harm to users also is required. The agency or program may employ additional criteria such as quality of development, user enthusiasm, conceptual soundness, etc.

3. Select products or practices for submission to the Panel on the basis of favorable prereviews.

4. Prepare (with the assistance, if necessary, of the producers, users, or evaluators of products or practices selected in step 3 above) submissions according to these guidelines and the attached forms. Since the written submissions are the principal basis of the Panel's judgments, it is important that this information be clear and complete. The required presubmission reviews for accuracy and completeness should have been completed by the program staff. The maximum permissible length of material to cover each submission is ten single-spaced typed pages including tables. The most important section, of course, is the summarization of the evidence of effectiveness. But the guidelines and form for submitting materials (appended) also require a brief but specific description of the product or practice and data on learner costs so that the Panel can consider the educational significance relative to the effort and costs required.

*As of January 1976, such products or practices must have been developed by federally supported programs, but may be copyrighted and published commercially. The Education Division Dissemination Policy Council, following comments from various groups, has decided that proprietary material not developed with federal funds will not be reviewed by the Panel.

5. Submit individual products and practices separately, if they are to be used separately, or if they are supported by different types of evidence.

6. Prepare a memorandum of transmittal signed by an OE Deputy Commissioner or an NIE Associate Director which states that the submission has been prereviewed and, to the best of his or her knowledge, is factually correct and that the products or practices are socially fair and apparently not harmful.

7. Submit 12 copies of the transmittal memorandum and each submission to the Recording Secretary of the Panel at the Office of the Assistant Secretary for Education.

8. Arrange for a representative of the OE or NIE program office making the submission to be present at the meeting. The representative should be qualified to answer any questions the Panel may have about the proposals. Program Offices may find it useful to have the assistance of evaluation, development, or user staff. The Panel encourages, but does not require, continuation of this practice. It should be emphasized that the submission is from the Education Division program office, not from the developer, evaluator, or user of the product or practice. The Division Officer is thus directly responsible for providing information on the submission at the Panel meeting.

B. JDRP Review Activities

1. The Chairperson of the Panel will schedule a meeting to review submissions. The date of the meeting will normally allow the Panel members five working days to review each submission.

2. All Panel meetings are open to the public.

3. Questions for clarification may be put to the Education Division program representatives, developers and evaluators by Panel members. A vote is taken on whether there is adequate evidence of effectiveness.

4. Seven members, including at least three members from each participating agency, constitute a quorum. (There are no substitutes.)

5. Voting will be by closed ballot, but the votes of the individual Panel members are announced by the Chairperson or the Recording Secretary after each vote has been completed.

6. Panel members disqualify themselves from voting on submissions made by programs in which they are directly involved.

7. A majority is required for a Panel decision that the evidence of effectiveness is convincing. A tie vote means a decision that the evidence is not convincing.

C. Follow-up Activities after Panel Review

1. Products or practices not favorably reviewed by the Panel may be resubmitted at any time.

2. The minutes from each meeting are distributed to all Panel members; OE Deputy Commissioners and NIE Associate Directors.

3. Agencies or programs are responsible for maintaining files or publishing catalogs of approved submissions. Currently a catalog of all JDRP-approved submissions is maintained by the U.S.O.E. Division of Educational Replication, Bureau of Elementary and Secondary Education.

4. Education Division Program Offices may endorse or support the dissemination of products or practices favorably reviewed by JDRP. Some offices do not have further reviews. Other offices require further product quality review before dissemination.

JDRP Criteria for Judging Effectiveness

The Joint Dissemination Review Panel has received requests for a statement of the "cutting points" on the criteria used in reviewing projects. The Panel has been unwilling to set forth precise specifications. The decision as to whether the evidence of effectiveness on any particular product or practice justifies use of Federal funds to promote dissemination is not automatic or mechanical. This decision involves judgment about the educational significance of the claimed effect and the credibility of the evidence supporting the claimed results. A decision that the evidence is not convincing is usually not due to a single fatal flaw, but to a failure of plausibility on several criteria.

The Panel believes that setting up highly specific standards relating to educational significance (e.g., the importance, number, magnitude, durability, and consistency of positive effects and their comparative advantage) and credibility of the supporting evidence (e.g., objectivity, reliability, validity, generalizability and causal nature of the data) would degrade rather than improve the review process. The Panel has felt, for example, that it would be irresponsible to assert that projects achieving a gain of 1.0 grade equivalent units should be accepted while a gain of 0.9 grade equivalent units should be rejected; or that a sample size of, say, 50 is acceptable, whereas a sample size of 49 is not.

It is because a complex judgment on plausibility is required -- one that takes into

account all of the evidence -- that the Joint Dissemination Review Panel is needed. Past experience indicates that leaving the judgment of what is acceptable up to each Education Division program produces variable and questionable results. On the other hand, developing a manual of boiler-plate standards would entail the arbitrariness cited above.

Some Guidance

All this may be true, but acknowledging it may not provide the guidance which Education Division program staffs need in order to identify products or practices which have a high likelihood of being approved.

The purpose of this section is to set forth in a general way, but in greater detail than is furnished in the Form for Submitting Materials to the Joint Dissemination Review Panel, the kinds of criteria the Panel uses in assessing the effectiveness of the products or practices submitted to it. This should provide program staffs with better guidance for selecting potential exemplary materials and methods, and for preparing submissions to the Panel.

Many of the products and practices submitted to the Panel have been in the areas of compensatory education, environmental education, reading, mathematics, bilingual education, and other similar areas of cognitive achievement or affective gain in individual elementary and secondary students. The criteria discussed below as illustrative relate mainly to these kinds of submissions.

Perhaps the best way of making clear the kinds of criteria the Panel uses in reaching its judgments on effectiveness is to set forth a series of general, common-sense questions. The questions focus on three issues: (1) Is there evidence that anything important happened that is consistent with the stated claims? (2) Is there evidence

that what happened is generalizable? and (3) Can this credibly be attributed to the product or practice?

1. How well did the children perform before they were subject to the product or practice in question, and how well did they perform after experiencing it? As an example, are there measures taken before and after the children experienced the innovation which show that they gained or improved?

2. Are the gains statistically reliable?

3. If there is evidence of change or improvement, is the gain large enough to be meaningful? That is, is it large enough to be both statistically reliable and educationally significant? Many submissions submitted to the panel do show apparent before-after mean changes in the expected direction, but they are either not statistically reliable or, even if statistically reliable, are too small in size to constitute meaningful educational improvement judged against commonly available experience.

4. Is there evidence that the improvements registered by the product or practice are stable and generalizable to other school settings? Or does it appear that they may be the result of special characteristics of the particular group of students who happened to use the product or practice when it was evaluated, or the unusual competence of the teacher who used it? These considerations are acute when the number of children is small or where there is only a single use of the product, e.g., where the evidence is based on one teacher and one classroom. Thus, it is desirable, when the number of cases is small, for there to have been multiple replications of the practice (e.g., different classrooms, different teachers, different schools, or different years).

5. Once it has been established that there have been gains, that they are large enough to be statistically and educationally meaningful, and that they are generalizable beyond the originating site, the central question remains: Can we be reasonably confident that the gains can be attributed to the innovation described in the submission rather than to practice on the measures, to normal maturation, to the regular education process, or to other special factors which may be impacting on the students? In order to provide some kind of answer to this crucially important question, submissions to the Panel should include some measure or estimate of what would have happened to the children if they had not been using the product or practice. The best way to do this, of course, is through a control group, but the Panel has approved many projects which did not contain a control group because the evidence of effectiveness contained some other persuasive measure or estimate of the effects of non-treatment. For example, comparisons of recent gains with previous year's scores or national norms often provide very persuasive evidence that the unusually large gains recorded can reasonably be attributed to the product or practice. Another example is time series comparisons against a baseline time period. In many cases, however, simple before-after data on the user groups are likely to be ambiguous when there are no parallel data on a comparable, untreated control group.

   The idea is to provide evidence ruling out rival hypotheses (factors which may be responsible for producing the effects attributed to the product or practice). Examples of rival hypotheses are:

* maturation -- psychological or physiological changes in the users.

* history -- events or administrative procedures which occurred in addition to the product or practice.

* selection/differential attrition -- when comparison groups are unequal on the relevant measures at the beginning (due to selection) or at the end (due to differential attrition).

* statistical regression -- e.g., when groups which are selected for their initial extreme scores on a given measure fall closer to their population mean score on retesting with the same measure.

* test effects -- increases in test scores due to the learning that results from frequent use of the same test, or to teaching to the test.

These are the main criteria by which the Panel tries to assess the evidence of effectiveness presented for products or practices. It is not possible to specify the precise point on any of these criteria which categorically differentiates acceptable from unacceptable submissions. Using all these criteria and examining all the evidence, the Panel tries to render judgments which, on one hand, will screen out those submissions whose evidence of effectiveness is not strong enough to support a federally funded dissemination effort, and on the other, will insure that those products and practices where there is such evidence, get moved into a wider process of dissemination which ultimately can improve education.

Other Kinds of Outcomes

It was noted above that these illustrations apply most appropriately to innovations directed at improving the achievement of elementary school and secondary school children. The illustrations discussed above would in many cases be quite inappropriate. The Education Division carries out a much wider variety of educational activities. Some activities seek to change attitudes or impart information. Others are aimed at bringing about changes in institutional procedures, at persuading state and local governments to adopt certain techniques, or at altering the ways in which state or local monies are spent. Still others seek to expand the pool of competent educational

practitioners through innovations in inservice or preservice training.

In all these cases, quite different outcome measures and evidence of effectiveness will be required. However, the same general logic implicit in the questions listed above apply to these products and practices as well: Is there persuasive, objective evidence that things are different after using the product or practice than they were before? And can we be reasonably confident that any such changes or improvements noted can be attributed to the effects of the item described rather than to something else?

Prereview Concerns

JDRP's primary responsibility is to judge the effectiveness of submitted products or practices. But prereviews before presentations to the Panel also serve as an additional opportunity to assure that other screening criteria which are the responsibilities of the Education Division's programs are met. As discussed in the first two sections of this paper, the Panel is concerned that the products or practices whose effectiveness is to be reviewed are clearly identified, that common implementation problems are described, and that full start-up and operating costs are presented so that schools interested in selecting the item will do so without misunderstanding its costs. Other federal concerns include social fairness and possible harm to users. All of these must be prereviewed by program offices.

Many other concerns for "good" products have been suggested. These include the conceptual soundness of the product or practice, the quality of the developmental process, acceptance by educators, the uniqueness of the submissions, how professional the materials look, and enthusiasm of parents and students.

The Panel regards these considerations as important. Program offices are encouraged to prereview project submissions to the Panel on all such criteria. Nonetheless, the Panel's sole function is to review only one category of criteria that may be considered

before dissemination: evidence of effectiveness. This may be considered a fairly large task by itself. Adopters should recognize, however, that JDRP only decides whether persuasive evidence of effectiveness consistent with the claims and goals has been presented. The product does not have a "universal seal of approval" from the federal government: that is not the Panel's function.

SUGGESTIONS FOR FURTHER READING

Tallmadge, G. Kasten; Horst, Donald P.; and Wood, Christine T. A Practical Guide to Measuring Project Impact on Student Achievement. Superintendent of Documents, Government Printing Office, Washington, D.C. Stock Number 017-080-0146.

Tallmadge, G. Kasten; and Horst, Donald P. A Procedural Guide for Validating Achievement Gains in Educational Projects. Superintendent of Documents, Government Printing Office, Washington, D.C.

Tallmadge, G. Kasten. The Joint Dissemination Review Panel Ideabook. Washington, D.C.: U.S. Office of Education/National Institute of Education, 1977 (U.S.G.P.O.)

FORM FOR SUBMITTING MATERIALS TO THE
JOINT DISSEMINATION REVIEW PANEL

PROGRAM AREA: (e.g., Title III, reading, career education, environmental education, education for the handicapped)

I. TITLE OF PRODUCT OR PRACTICE TO BE REVIEWED:

II. WHO DEVELOPED THE PRODUCT OR PRACTICE:

III. SOURCE AND LEVEL OF FUNDING:

IV. YEARS DURING WHICH THE PRODUCT OR PRACTICE WAS BEING DEVELOPED OR TESTED:

V. BRIEF DESCRIPTION OF THE PRODUCT OR PRACTICE FOR WHICH CLAIMS OF EFFECTIVENESS ARE BEING MADE AND TPE DATA:

1. What is the product or practice?

2. What claims of effectiveness are being made?

3. Who are the intended users and beneficiaries?

4. Characteristics of the groups on which the product or practice was developed and tested.

5. Costs for adoption of the product or practice (See table on p.2)

6. Interpretability of the quantitative measures: Evidence that the measures are reliable and valid as indicators of the effects claimed.

7. Credibility of evidence: Who collected and analyzed the data? What assurances are there that the findings are objective?

VI. EVIDENCE OF EFFECTIVENESS:

1. Evidence that something happened: What is the evidence that something happened?

2. Evidence of statistical reliability of the effects: What is the evidence that whatever happened, happened often enough and with sufficient reliability to be likely to happen again under similar circumstances?

3. Evidence that the effects are educationally meaningful: What is the evidence that the effects are large enough, powerful enough, or important enough to be educationally meaningful, regardless of the statistical reliability of the effects?

4. Evidence of generalizability to the populations for which the product or practice is intended: Evidence that the product or practice has been tested widely enough and under sufficiently diverse circumstances to give assurance that the effects claimed may be similar when the product or practice is used elsewhere for the populations intended.

5. Evidence that the effects are attributable to the product or practice and that rival hypotheses (such as maturation, practice effects, etc.) are ruled out: In order to be convincing, the evaluation need not be a strict experimental design, although this type of evidence is desirable. However, there must be some kind of high quality, objective, methodologically sound, quantitative evidence demonstrating that the product and practice in question is effective and superior to other more commonly used methods or approaches, and that the effects are caused by the product or practice.

VII. COST INFORMATION PER LEARNER TABLE SHELL

	INSTALLATION Non-recurring Costs	SUBSEQUENT YEARS (Recurring Costs)
Staff Training		
Special Facilities		
Equipment		
Consumables		
Other Costs		
TOTAL COSTS		

NOTE: This outline is to be used for the whole ten-page submission.

TRANSMITTAL MEMO FORM

TO: _____ Chairperson, JDRP

THROUGH (NIE Associate Director/USOE Assistant Commissioner):

_____

_____

_____

FROM: _____

SUBJECT: _____,

Submitted for Review for Joint Dissemination Review Panel

The attached submission is forwarded for review. The submission is factually accurate. I also affirm that the products or practices described in the submission have been favorably reviewed for social fairness (e.g., fairness with regard to sex, race, religion, or occupation) and have been examined with respect to possible harm to users. Limitations, if any, as to fairness or possible harm are noted in the submission.

Information on other prereview criteria applied by my office prior to JDRP submission is attached.

_____

_____

Upon completion of JDRP review, please send a copy of the minutes of the meeting to:

_____
Name

_____
Street

_____
City and State Zip Code

B. The External Implications of an Internal Review of Effectiveness: The DHEW Education Division's Joint Dissemination Review Panel

by Dr. Lois-ellin Datta*
National Institute of Education

In 1972, the Education Division of the U.S. Department of Health, Education and Welfare established a new policy on dissemination. Before the Division would take a position of advocacy in disseminating an educational product or practice for which claims of effectiveness were made, the evidence would have to be judged convincing by a panel of evaluators. A finding that the evidence was persuasive could open the door for either personal advocacy or Federal education dollars for dissemination. Without a favorable review, the product could not receive Education Division support for widespread use.

The purpose of this policy was two-fold. First, it would provide some Division-wide standard of evidence before a decision to disseminate was made, thus offering some Federal assurance that before public funds were used for dissemination, the product worked. Second, it would provide local consumers with more information about what the consequences of adopting or adapting the innovation would be.

The policy has been implemented though not without controversy (e.g., Hopkins, 1977) through the Joint Dissemination Review Panel (JDRP). Controversy has centered on four issues: the criteria by which effectiveness is judged; the review -- or not -- of proprietary products; equity of opportunity in providing convincing evidence when support for evaluation has varied, and the impact on evaluation practice and flow of products to the field as a result of JDRP's

*Presented at the April 4-8, 1977 American Educational Research Association meeting, New York, Division D Symposium with John Evans, Kasten Tallmadge, and Francis Thompson.

internal review functions.

The context of these controversies is the increase in attention to dissemination between 1972 and 1977. Federal funds for all dissemination and knowledge utilization over this six-year period increased in constant 1972 dollars and as a proportion of educational R & D funds. By 1975, of the $452,000,000 Federally obligated for educational knowledge production and utilization, 30% was spent in dissemination (Mason et al., 1977). Relative to evaluation, development, demonstrations, applied or basic research, dissemination is the growth stock for educational R & D.

Attention to dissemination has been spurred by public concern for direct benefits from educational R & D. This concern, expressed in hearings on research appropriations and in letters from constituents, has been translated into burning Congressional interest in getting products that work into the hands of teachers.

Among the approaches to dissemination taken by the U.S. Office of Education (USOE) and the National Institute of Education (NIE) are:

Continued support of the Educational Research Information Centers (ERIC initially funded in 1966), expansion of their activites from document collection to producing knowledge utilization reports, and expansion of the number of clearinghouses (NIE).

Development (1970) of the "It Works" series and Product Information Packaging (PIP's) system (USOE, 1972).

Creation within the Bureau of School Systems of USOE of a special office of dissemination and knowledge utilization (1974). This office administers the National-diffusion Network and the grant competitions for developers of JDRP-reviewed products (NIE).

Establishment of change agent networks to building state capacity to use educational research products in meeting state and local education needs (NIE, 1976).

Stimulation, through RFP's and grant competitions of front-end attention to research utilization for practice, legislation and policy. (USOE; NIE).

This incomplete summary adds up to a message: the days of writing journal articles and calling that dissemination are over.

The need for JDRP and the concerns about JDRP are related to the increase in funds for dissemination, and the development of dissemination as a new discipline. If the Federal Government puts 30% of all education R & D money into getting products that work into the hands of consumers, it better have some systematic ways of reviewing the evidence that these products are effective. The alternative is haphazard dissemination, susceptibility to pressures from developers (who are scarcely disinterested and objective) or placing the entire burden on consumers, most of whom cannot afford to support an educational effectiveness product review staff. The public appears to have had enough of educational products that don't work when adopted or still worse, may be harming rather than helping.

Controversy regarding JDRP does not appear to be with the principle of consumer protection and information about the innovation (such information to include but not be limited to claims of effectiveness). The problem is rather first in the location of the funnels between all ideas and some proven products (developer? state validation teams? Education Division? consumer?) when Federal dissemination funds are involved, and second in the acceptability of the criteria used for review (Crandall, 1975). Both of these issues

may reduce to whether (a) there is greater concern for ensuring that a product works as claimed even at the cost of screening out effective innovations where the evaluation has failed to capture these effects, versus (b) avoiding any loss in potentially effective innovations even at the cost of installing ineffective or harmful approaches. JDRP tries to achieve the first while minimizing loss of potentially effective innovations.

Conclusions about whether or not this internal review mechanism itself works depends on one's view of the criteria, the decision regarding proprietary products, and the issue of equity of opportunity in providing convincing evidence.

1. The Criteria: There are six criteria by which effectiveness is judged.

   The Panel asks:

   A. What is the evidence that something happened?

      Example: If the developer claims that the innovation improves high school completion rates, what are the rates now and are these an improvement over what they were before or what they otherwise would be, lacking the innovation?

   B. Is the evidence credible?

      Example: When teachers who developed an innovation and are implementing it create criterion-referenced tests which are not reviewed by subject matter authorities, and then administer these tests pre and post to their own students, the credibility of the evidence is low.

   C. Did something happen often enough that effects as large as those observed would be likely to happen again?

      Example: If every child who participated in the innovation gained more than every chile who did not in every class in every school for all the years of the study, it is plausible that the next child who participates will also gain more.

   D. Did something happen big enough that the effects are educationally significant?

      Example: If it costs $3,000 per child over and above regular expenses to raise the I.Q. of children whose initial I.Q. is 120 by ½ of an I.Q. point to 120.5, regardless of the statistical relia-

bility of the difference, the educational significance is dubious.

E. Did the something which happened often enough and big enough also occur under circumstances that are likely to be reproduceable?

Example: An approach to improving integration is tried out over a three-year period in school systems all over the country: urban, rural; large, small; rich, poor; high minority percent, low minority percent. The "installers" vary in age, sex and ethnicity but have in common having served as school board members and completed a one-month training course operated by the developer. It's plausible that other school systems would find their efforts at integration improved by adopting the approach as installed by ex-school board members who have passed the developer's training course.

F. Can what happened be attributed to the innovation?

Example: Improved use of existing resources to increase high school student access to career guidance is claimed. The measure is an increase from 20 percent of the senior high school students who said they had met with a career guidance counselor in the past six months to 90 percent. At the same time that the innovation intended to improve use of existing resources was used, the LEA doubled the number of counselors, the number of students decreased by 50%, the Department of Labor installed National Occupation Information System Terminals in local gas stations, pizza parlors, and record shops. Also, the Department of Commerce nationally broadcast career information spots and the NIE special career information TV program went on the air in the 6th through 9th grades of every junior high school.

The innovation may have been totally responsible for changes in high school students' reported access to career guidance counselors from 20% to 70%, but rival hypotheses have not been ruled out.

In framing these criteria, JDRP has tried first for a bare-bones minimum of what defines effectiveness and second to encourage judgment. The JDRP criteria are similar to those used by AIR in its study of exemplary programs. The novelty, if any, comes in the emphasis on persuasive evidence rather than yardsticks. The Panel seeks reasonable assurance on the criteria, not perfection. Innovations may be approved where evidence on one or another criterion is shaky, if as a whole the evidence seems convincing that the innovation is likely to benefit others if carefully adopted or adapted.

To some evaluators, JDRP standards are unrealistically high, particularly

in the areas of educational meaningfulness, attribution, and reproducibility. To other evaluators, JDRP standards are too low or not applied with sufficient rigor, particularly in the areas of what happened, statistical reliability and credibility of the data. Probably the best way to assess the assessor is to attend JDRP sessions, which are open to the public; to read the detailed notes of sessions since January 1976; or to role play review of submissions.

2. Review of Proprietary Products: As Fletcher (1976) has pointed out, JDRP does not review proprietary products unless these are a central part of an innovation developed with Federal Funds or for which a special position of advocacy in dissemination is sought. Responses to a letter of inquiry from publishers suggested a profound and what may be disquieting distrust of evaluators. Some publishers did not want JDRP review on ideological grounds. They felt the Federal government should not review the quality of educational products developed through the free enterprise system where use is constitutionally a state responsibility. Other publishers objected on practical grounds. They did not believe any evaluations were reliable enough, sensitive enough, or available soon enough to be gatekeepers on marketing or judgments regarding adoption.

From some perspectives, this may be a realistic assessment of the state of the art, and indeed, some evaluators are ready to conclude that evaluations only should be used formatively (Rayder 1977). From other perspectives, evaluations adequate to inform go/no go decisions are well within the state of the art technically, and failures to apply this technical expertise rest in politics, not better algebras for regression. JDRP is based on this latter assumption; nonetheless, members are often in the forefront of groups

examining this premise as well as working to strengthen the technology.

One consequence of the decision not to review proprietary materials is perhaps a greater equity for thin market distribution. Developers whose innovations are attractive to commercial distributors appear to go this route because of the marketing support. Proprietary materials typically are "thick market" with mass appeal. The JDRP route offers developers who have received Federal Support a dissemination system for "thin" market products. The two systems do not seem at present to compete, because they tend to disseminate different types of products.

It may be true, however, that later proprietary distribution of a product favorably reviewed by JDRP could give a competitive edge to an innovation. If so, and if consumers become educated to be concerned with evidence of effectiveness, the proprietary groups might well consider establishing their own procedures for independent, objective review of this evidence.

3. Equity of Opportunity: Many of the innovations submitted to JDRP were begun in 1972-73 through USOE demonstration awards. Evaluation in USOE is a centralized function for programs as a whole (such as Title I of ESEA), not as a service to identify exemplary instances. There are no funds for evaluation in some program awards and there are few evaluation staff in the Federal program management offices or in the regional offices. Yet the belief is strong that at least some of these individual LEA/SEA programs may be innovative and effective.

In contrast, through both USOE and NIE, some experimental and some development programs are funded amply for both formative and summative evaluations. Almost all NIE project officers have some research skills and all have access

to technical assistance on the evaluations.

Congressional and administration pressures to disseminate exemplary programs fall equally hard on the "have nots" and the "haves". The JDRP criteria are applied to submissions without regard to their source. However, the concern is rising that requiring documentation of effectiveness requires lead time, resources, and technical assistance which are unequally available.

Three actions already underway may reduce this concern:

1. Lead time: JDRP has been using many channels to ensure that developers, evaluators and disseminators know about the JDRP policy (e.g., Szigethy, 1977; Datta, 1977).

   It is probably fair to say that most evaluators and disseminators now know about JDRP. Second, NIE and some officers of USOE now routinely inform all researchers and developers at the time of the award of the JDRP policy. All NIE project officers and most USOE personnel who are judged to have a need to know have been oriented to JDRP.

2. Resources: Program funds can be set aside for evaluation under most titles. In the past, these frequently have purchased audits, process reports, consumer satisfaction surveys, and, less frequently, outcome evaluations. Why more money hasn't gone for outcome evaluations and why the quality of outcome evaluations when they are funded is often so low is unclear. One unkind interpretation is that there are a lot of incompetent evaluators in the business and that evaluation standards, such as those being developed under the leadership of Stufflebeam, are essential for consumer protection.

   A more charitable interpretation is that most evaluators are convinced that outcome-oriented evaluation should not be attempted until the innovation has stabilized and excellence of process achieved, but time runs out. Another reason might be that the politically weak position of most evaluators as employees of the program director gives evaluators almost no leverage for insisting on the essential elements of an outcome-oriented evaluation.

   It is also true in some instances that resources have been limited, particularly for the field-testing and multiple-year studies necessary for assessing reproducibility of effects. Increased awareness of the time and resources required may help provide the funds needed. In the meantime, evaluators could contribute by resisting the "having it both ways" claims that an innovation works so money should flow for dissemination while at the same time bemoaning lack of money to evaluate the innovation's effectiveness.

3. Technical Assistance: The technical assistance centers required for ESEA by the Education Amendments of 1974 may become models for other legislation. The centers are installing measures and designs that could let superintendents and school boards know what effects they're getting for the ESEA extra money and also permit cumulation of findings for reports to Congress. The Vocational Education Amendments of 1976 go a step further by specifying the criteria of effectiveness to be reported annually: student continuation in educational programs, placement in training related occupations, and employer satisfaction. The pull of such legislation coupled with the push of the Stufflebeam and JDRP evaluation standards, technical assistance centers, and a larger cadre of trained evaluators should make it possible for almost everyone to have access to evaluation assistance. Also, through NIE and USOE support, work is underway to improve measures and methodology in many outcome areas.

4. The Impact of JDRP on the Flow of Exemplary Practices to the Field:

JDRP reviews about 100 submissions yearly and recommends about 60% for dissemination. The Panel has been able to keep up with submissions from the program offices. Three submissions can be considered in a three hour meeting. At an average of two meetings a month, JDRP can accommodate about 75-100 submissions yearly as it is presently organized. There is no backlog of submissions, and the turnabout is usually three weeks. In this sense of affecting the process, preparing a JDRP submission and the review itself probably adds 2% to the total time in an average four year development cycle.

The concern, however, is less for the review time and more for whether JDRP standards are blocking exemplary projects from reaching the public, while ineffective or less exemplary innovations are disseminated through non-JDRP routes. The question is moot. In the absence of reasonably convincing evidence, there is no way to disentangle in the not-approved 40% an effective project to which an evaluation has been insensitive from an innovation that has little going for it but developer enthusiasm and consumer satisfaction. The argument has been made that if consumers like it, it has to be effective, but there is no evidence suggesting that people like what is good for them or that only effective programs are popular.

Perhaps "only" a 60% rate of favorable reviews seems too low, or too few projects even get through pre-reviews held in some Education Division program offices prior to submission. If so, the proper response may be to upgrade program and evaluation quality rather than to lower standards for judging effectiveness. It seems more likely that if JDRP can be faulted, it might be for leaning too strongly on an overall sense of the evidence, compensating for weaknesses on a single criterion where the Panel should insist on adequate evidence for all criteria.

In considering the external impact of this internal review of effectiveness, there seem to be at least three consequences which transcend the Education Division's management policies. These are:

1. The JDRP standards of effectiveness are likely to influence whatever criteria SEAs and LEAs may independently adopt. Judging from the IVD guidance, these SEA criteria are likely to go beyond JDRP in looking even more extensively into program development considerations prior to adoption and are as rigorous with regard to effectiveness, but seem to be weighted in a way that could approve an innovation failing to meet JDRP criteria because of high ratings on other criteria (USOE, undated). LEA's may prefer also the additional security of a Federal assurance of effectiveness to those provided by SEA review, especially in cross-state adoptions. At the least, JDRP may stimulate debate, consumer information, and use of criteria of effectiveness.

2. Thick market products developed with Federal funds may require more attention to outcome evaluation. If there is not a spillover effect to proprietary materials, paradoxically, JDRP may mean products affecting a fairly small number of students will be better tested for effectiveness than products affecting many learners.

3. The shape and course of educational innovation will be influenced increasingly by the resolution of tensions among three groups which do not have entirely common interests: developers whose interests tend toward maximum flexibility in rearranging ideas and simultaneously spreading the innovation everywhere; outcome effectiveness evaluators whose interests tend toward stability and controlled replication; and disseminators whose interests tend toward sufficient definition of what is essential in an innovation to permit reproducibility and maximum choice for consumers among a wide range of products.

JDRP-like mechanisms may create a common interest among the three groups when power shifts toward the outcome evaluator regardless of organ--izational position. If the balance swings too far, however, the result could be a premature "freezing" of the innovations before they're well developed or implemented. If the balance swings too little, the spread of educational innovations that apparently don't work will probably continue. If the balance is right, however, all consumers may have better information on impact and greater assurance of effectiveness for innovations.

This is a time, however, when it is easy to feel reviewed half to death. Our means of achieving implementation of principles seems too often based on distrust, suspicion and sanctions rather than respect, belief, and incentives. The need for any review group ought to be a case of being guilty until proven innocent, and the functioning of that group a matter of public scrutiny.

In the case of JDRP, it seems to me that our country is renowned for testing washing machines, aspirin, lipstick and asphalt before they're ready for consumer use. Providing assurance that proposed educational changes intended to influence the shape of individual lives and of societal futures probably have the beneficial effects claimed with no apparent harm seems the least we can do.

References

Crandall, David V. Program Validation: Four Case Studies. A Brief Report on Four Projects and Their Experience Before the Joint Dissemination Review Panel. Merrimac, Mass The Network if Innovation Schools, 1975.

Datta, Lois-ellin "There's more to effectiveness than achievement and some other questions and answers about the Joint Dissemination Review Panel." AERA/SIG Educational Research and Development Evaluators Newsletter. 1977, in press

Fletcher, Jerry L. "Notes from Washington." AERA/SIG Educational Research and Development Evaluators Newsletter, Vol. 5, No. 1, August 1, 1976

Hopkins, John E. "The JDRP: A Source of concern." AERA/SIG Educational Research and Development Evaluators Newsletter, February 1, 1977, 3-6

Mason, Ward S., Nelson, Carnot E., and Sowers, William M. Federal Funding for education knowledge production and utilization. Washington: National Institute of Education, 1977

Szigethy, Neil M. "Federal review panel sets new guidelines." CEDR Quarterly. 10, 1. Spring 1977, 20-21

U.S. Office of Education. Sharing Educational Success: A Handbook for Validation of Educational Practices. Washington: U.S. Office of Education, undated.

C. Comments on Dr. Lois-ellin Datta's Paper

by G. K. Tallmadge
RMC Research Corporation

The comments that I will make regarding Dr. Datta's presentation and the role and functioning of the JDRP will not be unbiased. They will, for the most part, represent the viewpoint of a hard-nosed outcome evaluator. Before proceeding, however, I would like to try to put this viewpoint into some kind of perspective.

I see the JDRP's job as both essential and impossible. It is essential because, for both moral and political reasons, Federal monies should not be spent disseminating as effective, educational products and practices that do not have beneficial impacts on their recipients. Such products and practices must be screened out. The job is impossible because, for both political and technical reasons, the information available to the panel members is never sufficient to remove all doubts about project effectiveness. Decisions must be based on incomplete evidence -- often on evidence known to be biased as a result of methodological errors in the evaluation itself.

It seems important from the onset to acknowledge that the JDRP will make two types of errors in carrying out their assigned task. They will approve products and practices that are really not effective, and they will disapprove products and practices that really are effective. Unfortunately, these two types of errors go hand in hand. Whenever the standards are set so as to minimize the probability that truly effective products and practices will be disapproved, the probability that truly ineffective projects will be approved automatically increases. The converse, of course, is also true.

While acknowledging that errors will be made, it is clearly appropriate to be concerned with the frequency of such errors and to seek to minimize their occurrence. To do this it is necessary to increase the discrimination power of the JDRP deliberation process. If it were possible to do this, the ratio of correct to incorrect decisions would be increased and the probability of occurrence of both types of errors would be correspondingly decreased.

There appear to be two avenues for increasing the power of the JDRP decision-making process. One is to increase the quality of the information available to the Panel and the other is to increase the proficiency with which the panel members assess the data. There is probably room for improvement in both areas, although it is hard for me to imagine where one could find a substantially better qualified group of 22 persons for the job. For this reason, I will confine my comments to the first area for potential improvement -- the data available to the JDRP.

I have read a substantial number of submissions made to the JDRP and I have read the minutes of the panel meetings -- including the questions posed by panel members to those making the submissions and the answers that were given. Many of the submissions reflect extensive and presumably expensive evaluations -- but very few of them reflect the state-of-the-art in evaluation technology. I would speculate that JDRP members often find themselves inclined to reject submissions because of inadequate evaluations yet feeling obligated to approve at least some proportion of them.

There <u>is</u> political pressure to approve submissions. There is additional political pressure not to let the evaluation tail wag the educational dog.

These are real forces that need to be reckoned with -- but it is important to note that they have the effect of causing the panel to approve products and practices in the absence of convincing proof of their benefits. Ineffective products may thus be disseminated with Federal dollars. They may be adopted with great expectations by districts that have been unsuccessful in their previous efforts to meet a particular educational need. Unfortunately, under these circumstances, the need will still not be met, the participating children will not benefit, credibility will be lost by all parties, and subsequent Federal efforts to aid education will be greeted with increased cynicism.

Clearly, all of these outcomes are undesirable. The only sure way to avoid them at present, however, would be to tighten up the JDRP approval criteria -- and then hardly any projects would be approved -- and the political pressures would increase. If the criteria were raised to the level where I, as a professional evaluator, would be satisfied and comfortable, the end result could well be the dissolution of the JDRP.

The question can be raised, "Are the evaluations really that bad?" I think the answer is that they are. I must admit that I have often been accused of having standards that are unrealistically high and I know that the JDRP has been similarly criticized. I think that such criticisms are unfounded. In many cases it is readily apparent from the submissions that more money has been spent doing a bad evaluation than would have been needed to do a good one.

It is also untrue that the JDRP imposes standards that cannot be met for practical reasons. The Panel has been accused of requiring control-group studies that cannot be implemented in real-world settings. No such requirements have

been imposed (although it would be less than candid to deny that evaluations that employ this type of design are usually very convincing). Evaluations using normative or historical data as a baseline for comparison can and have resulted in JDRP approval. But they must be done in careful and sensible ways. Such evaluations can be both meaningful and convincing if they are done correctly. No evaluation that is done without regard for the rules can be either meaningful or convincing.

If evaluations were better, the JDRP could do a better job and most of the criticisms directed at the group would disappear. It is probably safe to infer that current evaluations are so bad because, for years, the Federal government has insisted that Federally funded programs be evaluated but has failed to provide any guidelines as to how such evaluations should be conducted. Steps to correct this situation have recently been undertaken for the Title I program. As a result of that comprehensive effort entailing extensive model development, the preparation and dissemination of detailed documentation, and the establishment of regional technical assistance centers, significant improvements are to be expected in Title I evaluations. It is hoped that the increased awareness and understanding of evaluation issues will generalize to other programs.

As a direct attempt to improve the evaluations and other aspects of JDRP submissions, the Panel has recently undertaken the development of an Ideabook. This document will provide guidance to those preparing submissions to the JDRP. Each of the six criteria Dr. Datta described in her paper will be covered along with both positive and negative examples. The emphasis of the Ideabook will be a practical one stressing particularly ways to find and present convincing evidence without using random assignment, carefully controlled experiments, or standardized achievement tests.

The Ideabook and the new Title I evaluation and reporting system both represent efforts to improve the quality of locally conducted evaluations. They will almost certainly have some positive impact on relevant aspects of JDRP submissions. The improvements are likely to come slowly, however, and it seems clear that the JDRP cannot simply sit back and wait until more adequate evidence of effectiveness accompanies the submissions they must judge. At least for the present, then, I do not see any possibility for significant change in the role and methods of the Panel.

Personally, I might shoot for something closer to a 30% approval rate rather than the present 60% rate. But then it's easy for me to say that, since I am not on the receiving end of pressures from program staff personnel. I still feel, however, that rejecting possibly effective products and practices is the lesser evil when compared to approving possibly ineffective projects. Rejected products and practices can be resubmitted with additional evidence -- they are not shut out completely. On the other hand, once granted, an approval would be very difficult to withdraw.

I have only one suggestion to offer at this point, and I hesitate to offer it because it represents so radical a departure from current practice. Nevertheless, as an advocate of the hard-nosed evaluators, I suggest that JDRP approval not be regarded as a sufficient condition for the expenditure of Federal monies for dissemination. Rather I suggest that JDRP approval constitute a ticket for a thorough, Federally funded evaluation of the approved educational product or practice. Such an evaluation would obviously delay the dissemination of any truly effective products or practices, but it would also preclude the dissemination of those that are ineffective.